Guido Padilla 19200
Oscar Paredez 19109

1.

Dataset statistics

Number of variables	81
Number of observations	1460
Missing cells	6965
Missing cells (%)	5.9%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	924.0 KiB
Average record size in memory	648.1 B

Variable types

Numeric	30
Categorical	50
Boolean	1

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	PoolArea	PoolQC	Fence	MiscFeature	MiscVal	MoSold	YrSold	SaleType	SaleCond
0	1	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	2	2008	WD	N
1	2	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	5	2007	WD	N
2	3	60	RL	68.0	11250	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	NaN	0	9	2008	WD	N
3	4	70	RL	60.0	9550	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	NaN	0	2	2006	WD	Abr
4	5	60	RL	84.0	14260	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	NaN	0	12	2008	WD	N
5	6	50	RL	85.0	14115	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	MnPrv	Shed	700	10	2009	WD	N
6	7	20	RL	75.0	10084	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	8	2007	WD	N
7	8	60	RL	NaN	10382	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	Shed	350	11	2009	WD	N
8	9	50	RM	51.0	6120	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	4	2008	WD	Abr
9	10	190	RL	50.0	7420	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	1	2008	WD	N

	Id	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	...	WoodDeckSF	OpenPorchSF	EnclosedPorch	
count	1460.000000	1460.000000	1201.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1452.000000	1460.000000	...	1460.000000	1460.000000	1460.000000	1
mean	730.500000	56.897260	70.049958	10516.828082	6.099315	5.575342	1971.267808	1984.865753	103.685262	443.639726	...	94.244521	46.660274	21.954110	
std	421.610009	42.300571	24.284752	9981.264932	1.382997	1.112799	30.202904	20.645407	181.066207	456.098091	...	125.338794	66.256028	61.119149	
min	1.000000	20.000000	21.000000	1300.000000	1.000000	1.000000	1872.000000	1950.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	
25%	365.750000	20.000000	59.000000	7553.500000	5.000000	5.000000	1954.000000	1967.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	
50%	730.500000	50.000000	69.000000	9478.500000	6.000000	5.000000	1973.000000	1994.000000	0.000000	383.500000	...	0.000000	25.000000	0.000000	
75%	1095.250000	70.000000	80.000000	11601.500000	7.000000	6.000000	2000.000000	2004.000000	166.000000	712.250000	...	168.000000	68.000000	0.000000	
max	1460.000000	190.000000	313.000000	215245.000000	10.000000	9.000000	2010.000000	2010.000000	1600.000000	5644.000000	...	857.000000	547.000000	552.000000	

...	WoodDeckSF	OpenPorchSF	EnclosedPorch	3SsnPorch	ScreenPorch	PoolArea	MiscVal	MoSold	YrSold	SalePrice
...	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000
...	94.244521	46.660274	21.954110	3.409589	15.060959	2.758904	43.489041	6.321918	2007.815753	180921.195890
...	125.338794	66.256028	61.119149	29.317331	55.757415	40.177307	496.123024	2.703626	1.328095	79442.502883
...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	2006.000000	34900.000000
...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	5.000000	2007.000000	129975.000000
...	0.000000	25.000000	0.000000	0.000000	0.000000	0.000000	0.000000	6.000000	2008.000000	163000.000000
...	168.000000	68.000000	0.000000	0.000000	0.000000	0.000000	0.000000	8.000000	2009.000000	214000.000000
...	857.000000	547.000000	552.000000	508.000000	480.000000	738.000000	15500.000000	12.000000	2010.000000	755000.000000

2.

a. Cuantitativa
   i. Discreta
      1. Id
      2. MSSubClass
      3. LotFrontage
      4. LotArea
      5. YearBuilt
      6. YearRemodAdd
      7. MasVnrArea
      8. BsmtFinSF1
      9. BsmtFinSF2
      10. BsmtUnfSF
      11. TotalBsmtSF

12. 1stFlrSF
13. 2ndFlrSF
14. LowQualFinSF
15. GrLivArea
16. BsmtFullBath
17. BsmtHalfBath
18. FullBath
19. HalfBath
20. BedroomAbvGr
21. KitchenAbvGr
22. TotRmsAbvGrd
23. Fireplaces
24. GarageYrBlt
25. GarageCars
26. GarageArea
27. WoodDeckSF
28. OpenPorchSF
29. EnclosedPorch
30. 3SsnPorch
31. ScreenPorch
32. PoolArea
33. MiscVal
34. MoSold
35. YrSold

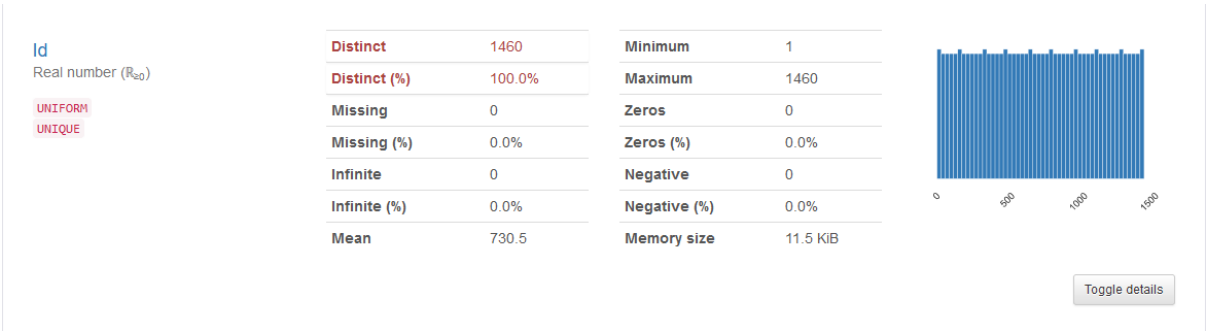
ii. Continua
    1.

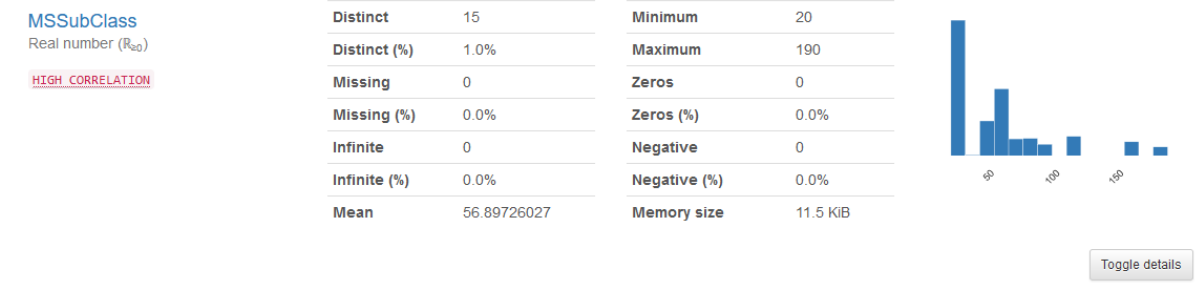
b. Cualitativa o Categórica
    i. Alley
    ii. Street
    iii. LotShape
    iv. LandContour
    v. Utilities
    vi. LotConfig
    vii. LandSlope
    viii. Neighborhood
    ix. Condition1
    x. Condition2
    xi. BldgType
    xii. HouseStyle
    xiii. OverallQual
    xiv. OverallCond
    xv. RoofStyle
    xvi. RoofMatl
    xvii. Exterior1st
    xviii. Exterior2nd
    xix. MasVnrType
    xx. ExterQual
    xxi. ExterCond

xxii. Foundation
xxiii. BsmtQual
xxiv. BsmtCond
xxv. BsmtExposure
xxvi. BsmtFinType1
xxvii. BsmtFinType2
xxviii. Heating
xxix. HeatingQC
xxx. CentralAir
xxxi. Electrical
xxxii. KitchenQual
xxxiii. Functional
xxxiv. FireplaceQu
xxxv. GarageType
xxxvi. GarageFinish
xxxvii. GarageQual
xxxviii. GarageCond
xxxix. PavedDrive
xl. PoolQC
xli. Fence
xlii. MiscFeature
xliii. SaleType
xliv. SaleCondition

3.

Variables cualitativas





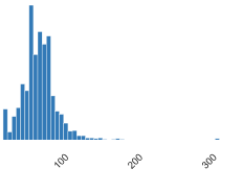
LotFrontage

Real number ($\mathbb{R}_{\geq 0}$)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
MISSING

Distinct	110
Distinct (%)	9.2%
Missing	259
Missing (%)	17.7%
Infinite	0
Infinite (%)	0.0%
Mean	70.04995837

Minimum	21
Maximum	313
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



Toggle details

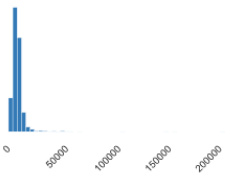
LotArea

Real number ($\mathbb{R}_{\geq 0}$)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION

Distinct	1073
Distinct (%)	73.5%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	10516.82808

Minimum	1300
Maximum	215245
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



Toggle details

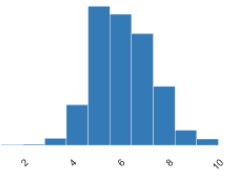
OverallQual

Real number ($\mathbb{R}_{\geq 0}$)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION

Distinct	10
Distinct (%)	0.7%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	6.099315068

Minimum	1
Maximum	10
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



Toggle details

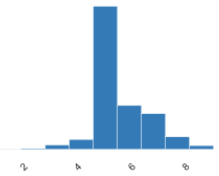
OverallCond

Real number ($\mathbb{R}_{\geq 0}$)

HIGH CORRELATION

Distinct	9
Distinct (%)	0.6%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	5.575342466

Minimum	1
Maximum	9
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



Toggle details

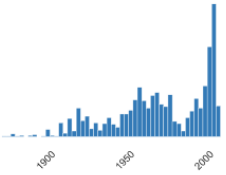
YearBuilt

Real number (ℝ≥0)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION

Distinct	112
Distinct (%)	7.7%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1971.267808

Minimum	1872
Maximum	2010
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



Toggle details

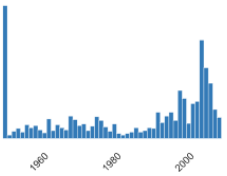
YearRemodAdd

Real number (ℝ≥0)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION

Distinct	61
Distinct (%)	4.2%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1984.865753

Minimum	1950
Maximum	2010
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



Toggle details

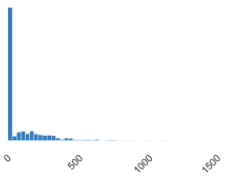
MasVnrArea

Real number (ℝ≥0)

HIGH CORRELATION
ZEROS

Distinct	327
Distinct (%)	22.5%
Missing	8
Missing (%)	0.5%
Infinite	0
Infinite (%)	0.0%
Mean	103.6852617

Minimum	0
Maximum	1600
Zeros	861
Zeros (%)	59.0%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



Toggle details

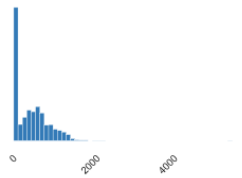
BsmtFinSF1

Real number (ℝ≥0)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
ZEROS

Distinct	637
Distinct (%)	43.6%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	443.639726

Minimum	0
Maximum	5644
Zeros	467
Zeros (%)	32.0%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



Toggle details

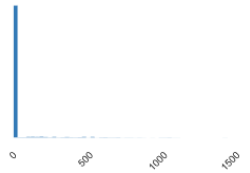
BsmtFinSF2

Real number ($\mathbb{R}_{\geq 0}$)

HIGH CORRELATION
ZEROS

Distinct	144
Distinct (%)	9.9%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	46.54931507

Minimum	0
Maximum	1474
Zeros	1293
Zeros (%)	88.6%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



Toggle details

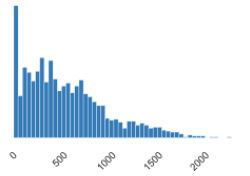
BsmtUnfSF

Real number ($\mathbb{R}_{\geq 0}$)

HIGH CORRELATION
HIGH CORRELATION
ZEROS

Distinct	780
Distinct (%)	53.4%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	567.240411

Minimum	0
Maximum	2336
Zeros	118
Zeros (%)	8.1%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



Toggle details

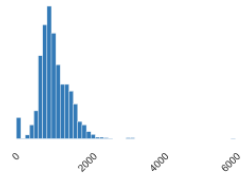
TotalBsmtSF

Real number ($\mathbb{R}_{\geq 0}$)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
ZEROS

Distinct	721
Distinct (%)	49.4%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1057.429452

Minimum	0
Maximum	6110
Zeros	37
Zeros (%)	2.5%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



Toggle details

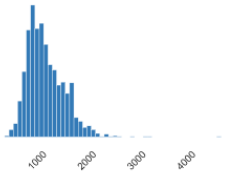
1stFlrSF

Real number ($\mathbb{R}_{\geq 0}$)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION

Distinct	753
Distinct (%)	51.6%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1162.626712

Minimum	334
Maximum	4692
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



Toggle details

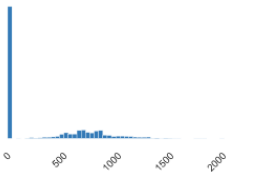
2ndFlrSF

Real number ($\mathbb{R}_{\geq 0}$)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
ZEROS

Distinct	417
Distinct (%)	28.6%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	346.9924658

Minimum	0
Maximum	2065
Zeros	829
Zeros (%)	56.8%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



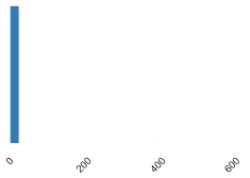
Toggle details

LowQualFinSF

Real number (ℝ≥0)

HIGH CORRELATION
ZEROS

Distinct	24	**Minimum**	0
Distinct (%)	1.6%	**Maximum**	572
Missing	0	**Zeros**	1434
Missing (%)	0.0%	**Zeros (%)**	98.2%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	5.844520548	**Memory size**	11.5 KiB



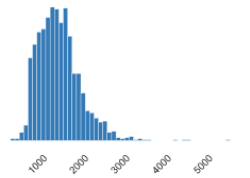
Toggle details

GrLivArea

Real number (ℝ≥0)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION

Distinct	861	**Minimum**	334
Distinct (%)	59.0%	**Maximum**	5642
Missing	0	**Zeros**	0
Missing (%)	0.0%	**Zeros (%)**	0.0%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	1515.463699	**Memory size**	11.5 KiB



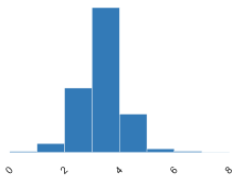
Toggle details

BedroomAbvGr

Real number (ℝ≥0)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION

Distinct	8	**Minimum**	0
Distinct (%)	0.5%	**Maximum**	8
Missing	0	**Zeros**	6
Missing (%)	0.0%	**Zeros (%)**	0.4%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	2.866438356	**Memory size**	11.5 KiB



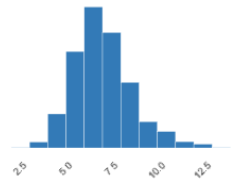
Toggle details

TotRmsAbvGrd

Real number (ℝ≥0)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION

Distinct	12	**Minimum**	2
Distinct (%)	0.8%	**Maximum**	14
Missing	0	**Zeros**	0
Missing (%)	0.0%	**Zeros (%)**	0.0%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	6.517808219	**Memory size**	11.5 KiB



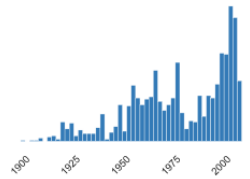
Toggle details

GarageYrBlt

Real number (ℝ≥0)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
MISSING

Distinct	97	**Minimum**	1900
Distinct (%)	7.0%	**Maximum**	2010
Missing	81	**Zeros**	0
Missing (%)	5.5%	**Zeros (%)**	0.0%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	1978.506164	**Memory size**	11.5 KiB



Toggle details

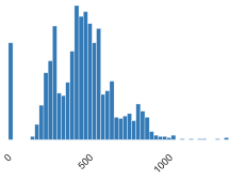
GarageArea

Real number (ℝ≥0)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
ZEROS

Distinct	441
Distinct (%)	30.2%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	472.980137

Minimum	0
Maximum	1418
Zeros	81
Zeros (%)	5.5%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



Toggle details

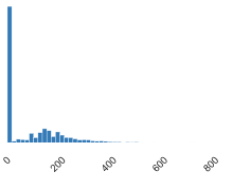
WoodDeckSF

Real number (ℝ≥0)

HIGH CORRELATION
ZEROS

Distinct	274
Distinct (%)	18.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	94.24452055

Minimum	0
Maximum	857
Zeros	761
Zeros (%)	52.1%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



Toggle details

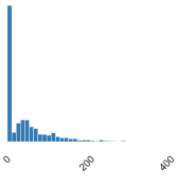
OpenPorchSF

Real number (ℝ≥0)

HIGH CORRELATION
ZEROS

Distinct	202
Distinct (%)	13.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	46.66027397

Minimum	0
Maximum	547
Zeros	656
Zeros (%)	44.9%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



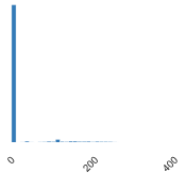
EnclosedPorch

Real number (ℝ≥0)

HIGH CORRELATION
ZEROS

Distinct	120
Distinct (%)	8.2%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	21.95410959

Minimum	0
Maximum	552
Zeros	1252
Zeros (%)	85.8%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



Toggle details

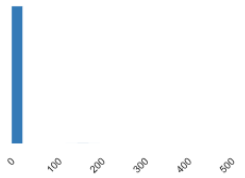
3SsnPorch

Real number (ℝ≥0)

ZEROS

Distinct	20
Distinct (%)	1.4%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	3.409580041

Minimum	0
Maximum	508
Zeros	1436
Zeros (%)	98.4%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB

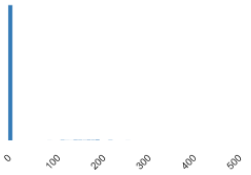


ScreenPorch

Real number (ℝ≥0)

ZEROS

Distinct	76	**Minimum**	0
Distinct (%)	5.2%	**Maximum**	480
Missing	0	**Zeros**	1344
Missing (%)	0.0%	**Zeros (%)**	92.1%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	15.0609589	**Memory size**	11.5 KiB



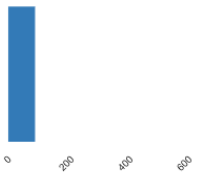
Toggle details

PoolArea

Real number (ℝ≥0)

HIGH CORRELATION
ZEROS

Distinct	8	**Minimum**	0
Distinct (%)	0.5%	**Maximum**	738
Missing	0	**Zeros**	1453
Missing (%)	0.0%	**Zeros (%)**	99.5%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	2.75890411	**Memory size**	11.5 KiB



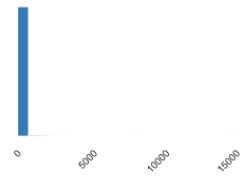
Toggle details

MiscVal

Real number (ℝ≥0)

HIGH CORRELATION
SKEWED
ZEROS

Distinct	21	**Minimum**	0
Distinct (%)	1.4%	**Maximum**	15500
Missing	0	**Zeros**	1408
Missing (%)	0.0%	**Zeros (%)**	96.4%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	43.4890411	**Memory size**	11.5 KiB

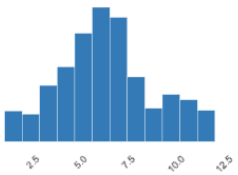


Toggle details

MoSold

Real number (ℝ≥0)

Distinct	12	**Minimum**	1
Distinct (%)	0.8%	**Maximum**	12
Missing	0	**Zeros**	0
Missing (%)	0.0%	**Zeros (%)**	0.0%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	6.321917808	**Memory size**	11.5 KiB



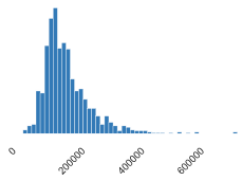
Toggle details

SalePrice

Real number (ℝ≥0)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION

Distinct	663	**Minimum**	34900
Distinct (%)	45.4%	**Maximum**	755000
Missing	0	**Zeros**	0
Missing (%)	0.0%	**Zeros (%)**	0.0%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	180921.1959	**Memory size**	11.5 KiB



Toggle details

Variables Cuantitativas

MSZoning

Categorical

HIGH CORRELATION

Distinct	5
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB

Toggle details

Street

Categorical

Distinct	2
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB

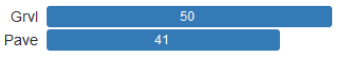
Toggle details

Alley

Categorical

HIGH CORRELATION
MISSING

Distinct	2
Distinct (%)	2.2%
Missing	1369
Missing (%)	93.8%
Memory size	11.5 KiB

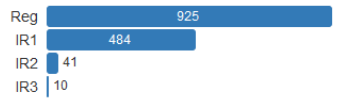


Toggle details

LotShape

Categorical

Distinct	4
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB



LandContour

Categorical

HIGH CORRELATION

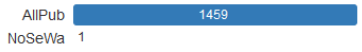
Distinct	4
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB

Toggle details

Utilities

Categorical

Distinct	2
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB



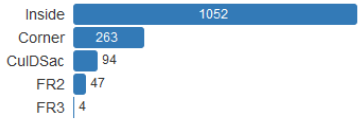
Toggle details

LotConfig

Categorical

HIGH CORRELATION

Distinct	5
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB



LandSlope
Categorical

HIGH CORRELATION

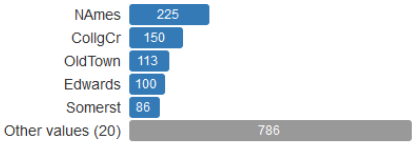
Distinct	3
Distinct (%)	0.2%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB

Toggle details

Neighborhood
Categorical

HIGH CORRELATION

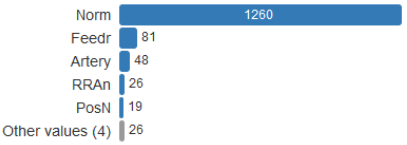
Distinct	25
Distinct (%)	1.7%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB



Toggle details

Condition1
Categorical

Distinct	9
Distinct (%)	0.6%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB

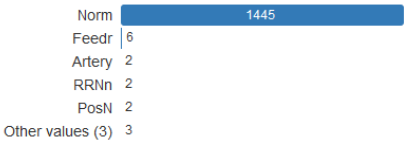


Toggle details

Condition2
Categorical

HIGH CORRELATION

Distinct	8
Distinct (%)	0.5%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB



Toggle details

BldgType
Categorical

HIGH CORRELATION

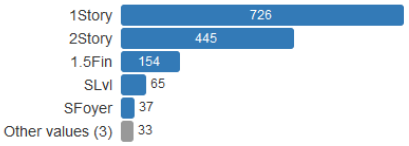
Distinct	5
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB

Toggle details

HouseStyle
Categorical

HIGH CORRELATION

Distinct	8
Distinct (%)	0.5%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB

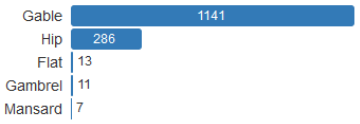


RoofStyle

Categorical

HIGH CORRELATION

Distinct	6
Distinct (%)	0.4%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB



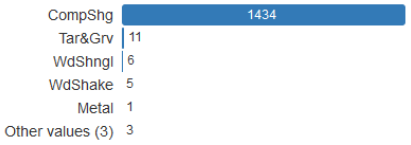
Toggle details

RoofMatl

Categorical

HIGH CORRELATION

Distinct	8
Distinct (%)	0.5%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB



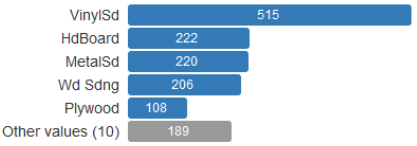
Toggle details

Exterior1st

Categorical

HIGH CORRELATION

Distinct	15
Distinct (%)	1.0%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB



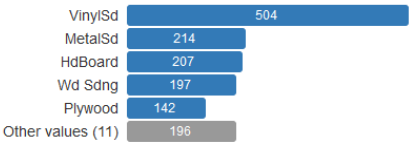
Toggle details

Exterior2nd

Categorical

HIGH CORRELATION

Distinct	16
Distinct (%)	1.1%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB



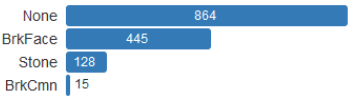
Toggle details

MasVnrType

Categorical

HIGH CORRELATION

Distinct	4
Distinct (%)	0.3%
Missing	8
Missing (%)	0.5%
Memory size	11.5 KiB



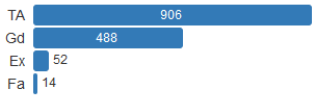
Toggle details

ExterQual

Categorical

HIGH CORRELATION

Distinct	4
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB



Toggle details

ExterCond

Categorical

HIGH CORRELATION

Distinct	5
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB

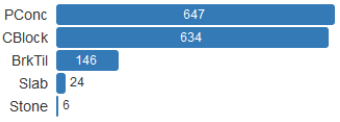
Toggle details

Foundation

Categorical

HIGH CORRELATION

Distinct	6
Distinct (%)	0.4%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB



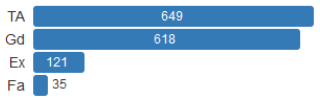
Toggle details

BsmtQual

Categorical

HIGH CORRELATION
MISSING

Distinct	4
Distinct (%)	0.3%
Missing	37
Missing (%)	2.5%
Memory size	11.5 KiB



Toggle details

BsmtCond

Categorical

HIGH CORRELATION
MISSING

Distinct	4
Distinct (%)	0.3%
Missing	37
Missing (%)	2.5%
Memory size	11.5 KiB

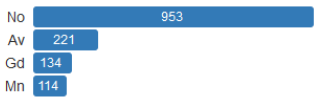
Toggle details

BsmtExposure

Categorical

HIGH CORRELATION
MISSING

Distinct	4
Distinct (%)	0.3%
Missing	38
Missing (%)	2.6%
Memory size	11.5 KiB



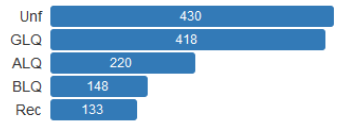
Toggle details

BsmtFinType1

Categorical

HIGH CORRELATION
MISSING

Distinct	6
Distinct (%)	0.4%
Missing	37
Missing (%)	2.5%
Memory size	11.5 KiB



Toggle details

BsmtFinType2

Categorical

HIGH CORRELATION
MISSING

Distinct	6
Distinct (%)	0.4%
Missing	38
Missing (%)	2.6%
Memory size	11.5 KiB

Toggle details

Heating

Categorical

HIGH CORRELATION

Distinct	6
Distinct (%)	0.4%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB

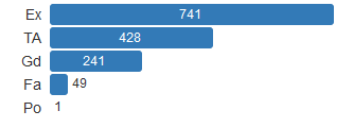
Toggle details

HeatingQC

Categorical

HIGH CORRELATION

Distinct	5
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB



Toggle details

CentralAir

Boolean

HIGH CORRELATION

Distinct	2
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.6 KiB

Electrical

Categorical

HIGH CORRELATION

Distinct	5
Distinct (%)	0.3%
Missing	1
Missing (%)	0.1%
Memory size	11.5 KiB

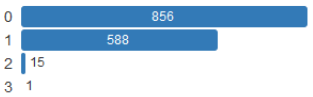
Toggle details

BsmtFullBath

Categorical

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION

Distinct	4
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB



Toggle details

BsmtHalfBath

Categorical

HIGH CORRELATION

Distinct	3
Distinct (%)	0.2%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB

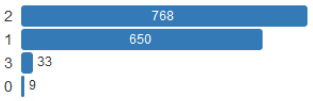
Toggle details

FullBath

Categorical

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION

Distinct	4
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB

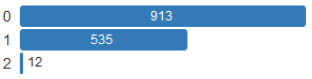


HalfBath

Categorical

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION

Distinct	3
Distinct (%)	0.2%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB



Toggle details

KitchenAbvGr

Categorical

HIGH CORRELATION

Distinct	4
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB

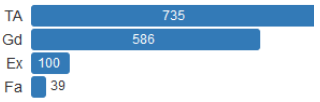
Toggle details

KitchenQual

Categorical

HIGH CORRELATION

Distinct	4
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB



Toggle details

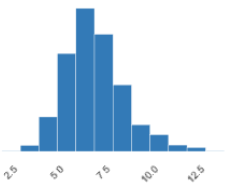
TotRmsAbvGrd

Real number ($\mathbb{R}_{\geq 0}$)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION

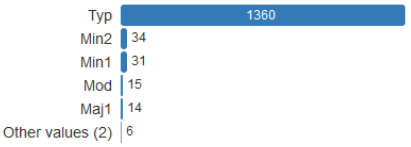
Distinct	12
Distinct (%)	0.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	6.517808219

Minimum	2
Maximum	14
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	11.5 KiB



Functional
Categorical

Distinct	7
Distinct (%)	0.5%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB

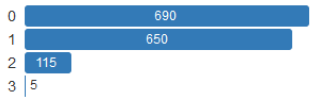


Toggle details

Fireplaces
Categorical

HIGH CORRELATION
HIGH CORRELATION

Distinct	4
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB

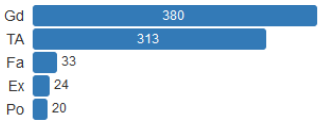


Toggle details

FireplaceQu
Categorical

HIGH CORRELATION
MISSING

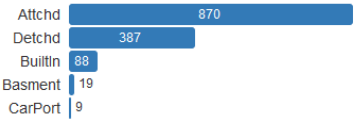
Distinct	5
Distinct (%)	0.6%
Missing	690
Missing (%)	47.3%
Memory size	11.5 KiB



GarageType
Categorical

HIGH CORRELATION
MISSING

Distinct	6
Distinct (%)	0.4%
Missing	81
Missing (%)	5.5%
Memory size	11.5 KiB

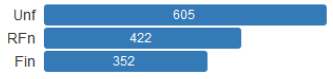


Toggle details

GarageFinish
Categorical

HIGH CORRELATION
MISSING

Distinct	3
Distinct (%)	0.2%
Missing	81
Missing (%)	5.5%
Memory size	11.5 KiB

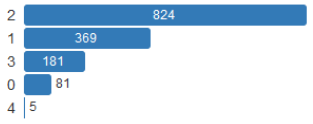


Toggle details

GarageCars
Categorical

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION

Distinct	5
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB



Toggle details

GarageQual
Categorical

HIGH CORRELATION
MISSING

Distinct	5
Distinct (%)	0.4%
Missing	81
Missing (%)	5.5%
Memory size	11.5 KiB

Toggle details

GarageCond
Categorical

HIGH CORRELATION
MISSING

Distinct	5
Distinct (%)	0.4%
Missing	81
Missing (%)	5.5%
Memory size	11.5 KiB

Toggle details

PavedDrive
Categorical

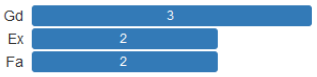
HIGH CORRELATION

Distinct	3
Distinct (%)	0.2%
Missing	0
Missing (%)	0.0%
Memory size	11.5 KiB

PoolQC
Categorical

HIGH CORRELATION
MISSING

Distinct	3
Distinct (%)	42.9%
Missing	1453
Missing (%)	99.5%
Memory size	11.5 KiB

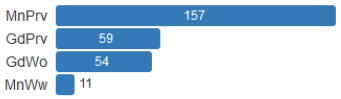


Toggle details

Fence
Categorical

MISSING

Distinct	4
Distinct (%)	1.4%
Missing	1179
Missing (%)	80.8%
Memory size	11.5 KiB



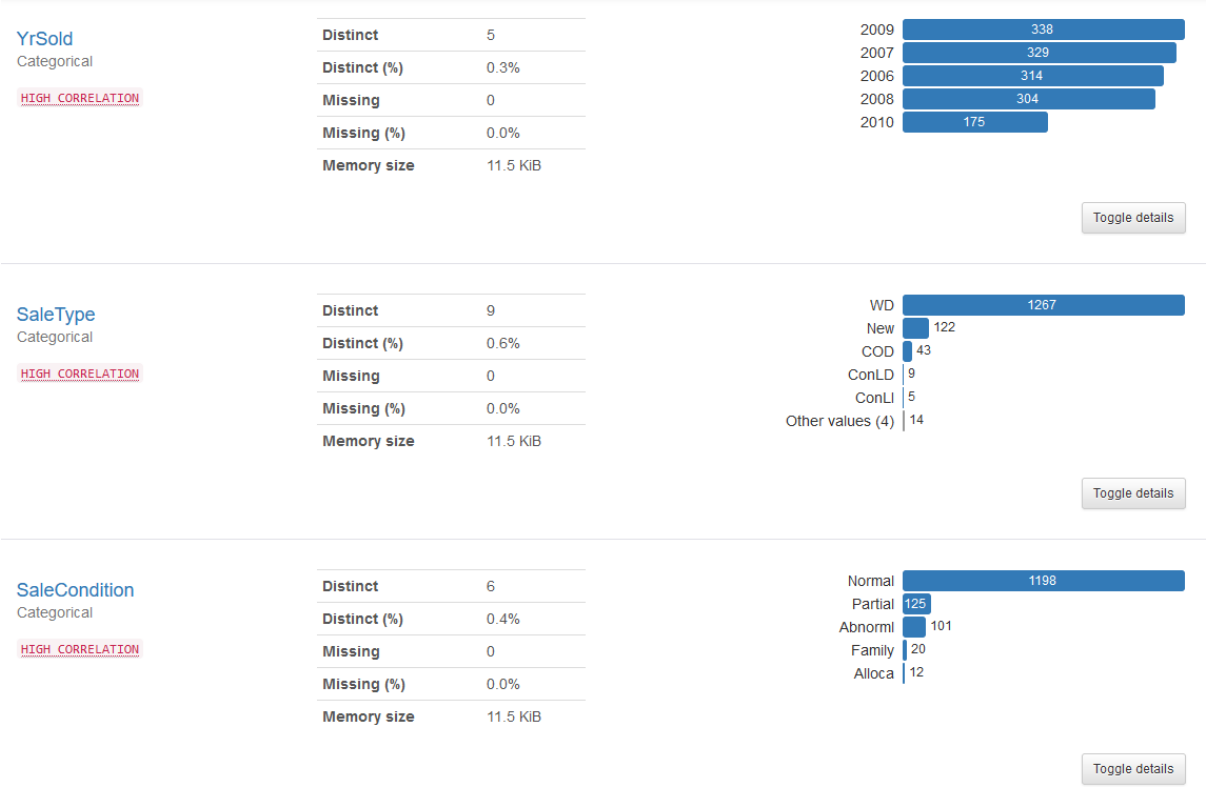
Toggle details

MiscFeature
Categorical

HIGH CORRELATION
MISSING

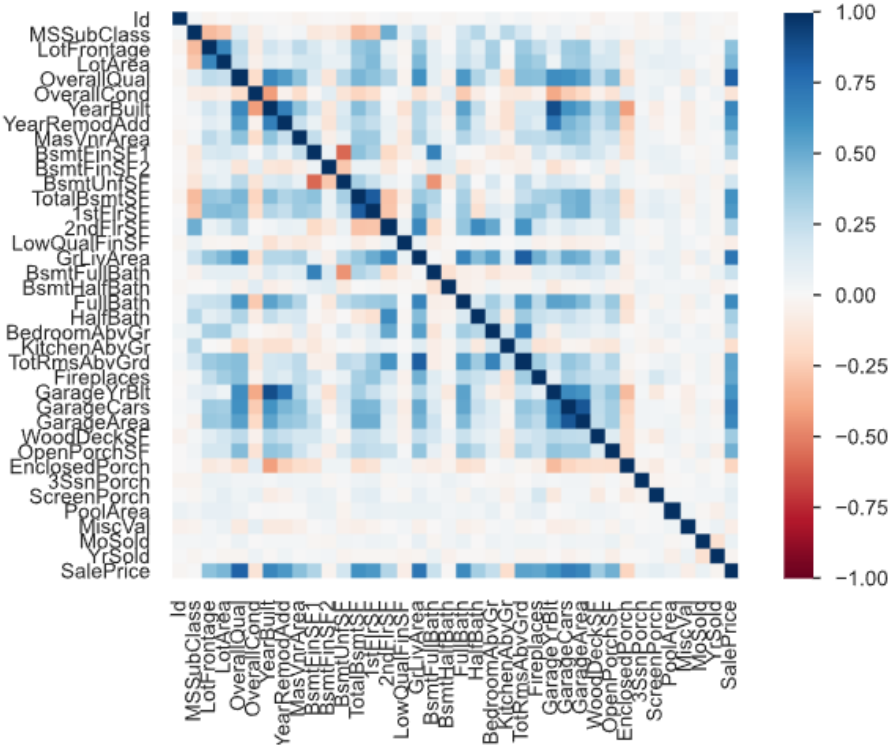
Distinct	4
Distinct (%)	7.4%
Missing	1406
Missing (%)	96.3%
Memory size	11.5 KiB

Toggle details



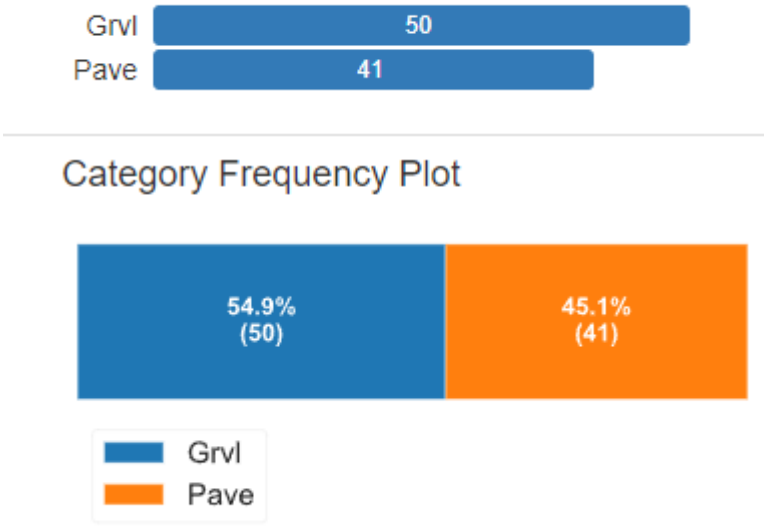
Para el análisis exploratorio se utilizó la herramienta de Profile Report para poder detallar cada una de las variables del dataframe. Se realizó la separación de variables cuantitativas y cualitativas y en base a eso se desarrollaron histogramas y gráficas de frecuencia.

4.

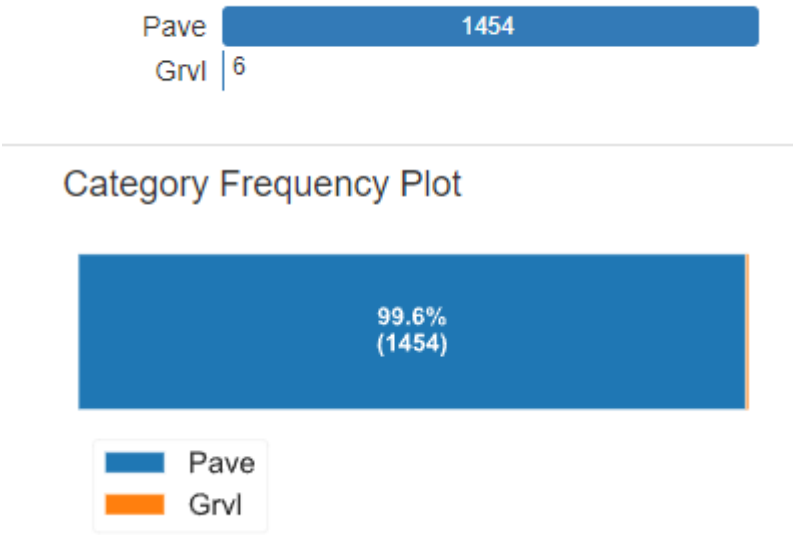


Al realizar la matriz de correlación pudimos notar ciertos puntos importantes, primero existen variables en el conjunto de datos que aportan muy poco al mismo, estás siendo aquellas que poseen un valor casi nulo en la misma matriz de correlación, esto ayudando a poder descartarlas para poseer resultados más certeros al momento de analizar y realizar test, segundo existen tanto relaciones directamente proporcionales como inversamente proporcionales lo cual será a tomar a consideración al realizar futuras pruebas.

5. Cualitativa o Categórica
    a. Alley



    b. Street



    c. LotShape
    d. LandContour
    e. Utilities
    f. LotConfig
    g. LandSlope

h. Neighborhood
i. Condition1
j. Condition2
k. BldgType
l. HouseStyle
m. OverallQual
n. OverallCond
o. RoofStyle
p. RoofMatl
q. Exterior1st
r. Exterior2nd
s. MasVnrType
t. ExterQual
u. ExterCond
v. Foundation
w. BsmtQual
x. BsmtCond
y. BsmtExposure
z. BsmtFinType1
aa. BsmtFinType2
bb. Heating
cc. HeatingQC
dd. CentralAir
ee. Electrical
ff. KitchenQual
gg. Functional
hh. FireplaceQu
ii. GarageType
jj. GarageFinish
kk. GarageQual
ll. GarageCond
mm. PavedDrive
nn. PoolQC
oo. Fence
pp. MiscFeature
qq. SaleType
rr. SaleCondition

6.

```
import pandas as pd
from factor_analyzer import FactorAnalyzer
import matplotlib.pyplot as plt
from factor_analyzer.factor_analyzer import calculate_bartlett_sphericity
data_test = data[cuantitativas]
data_test.dropna(inplace=True)
chi_square_value,p_value=calculate_bartlett_sphericity(data_test)
chi_square_value, p_value
```

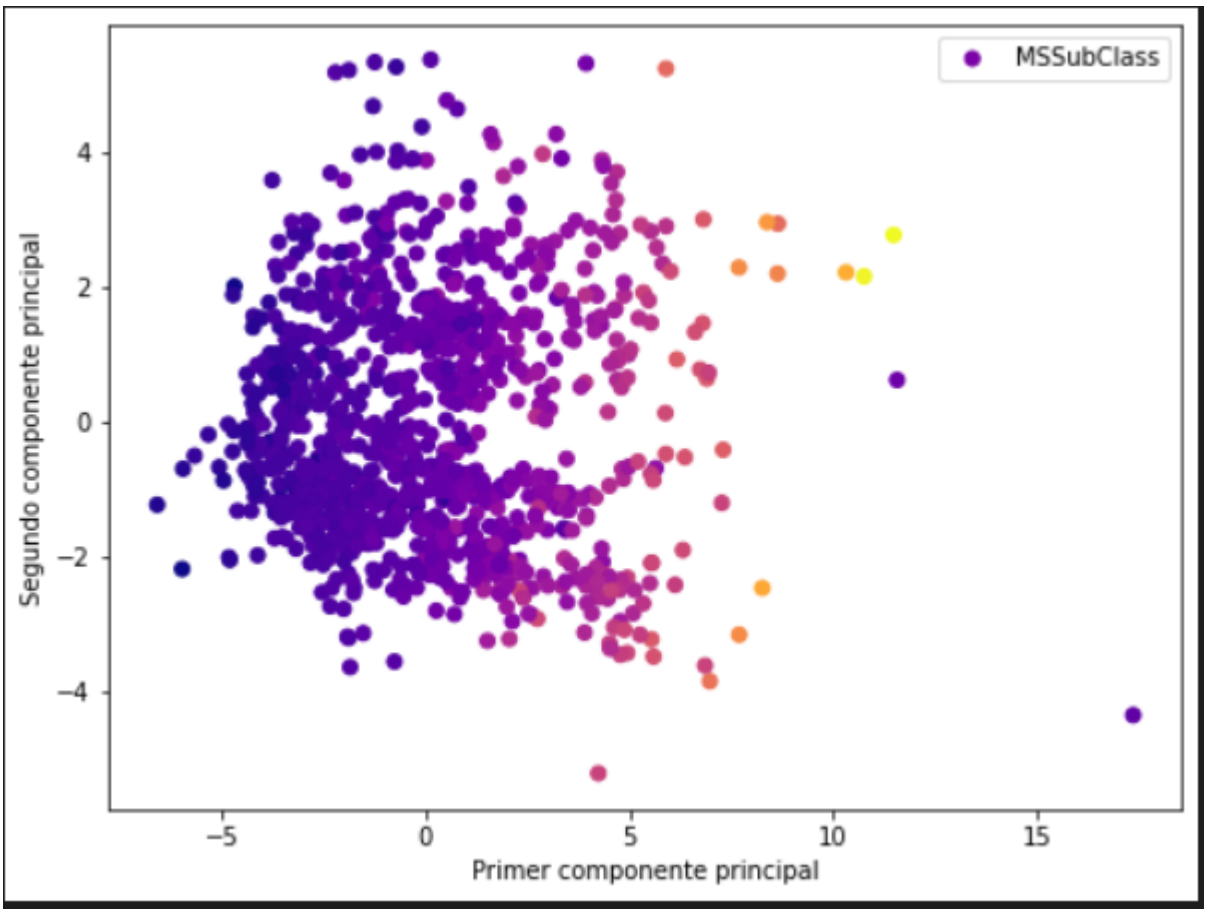
✓ 0.6s

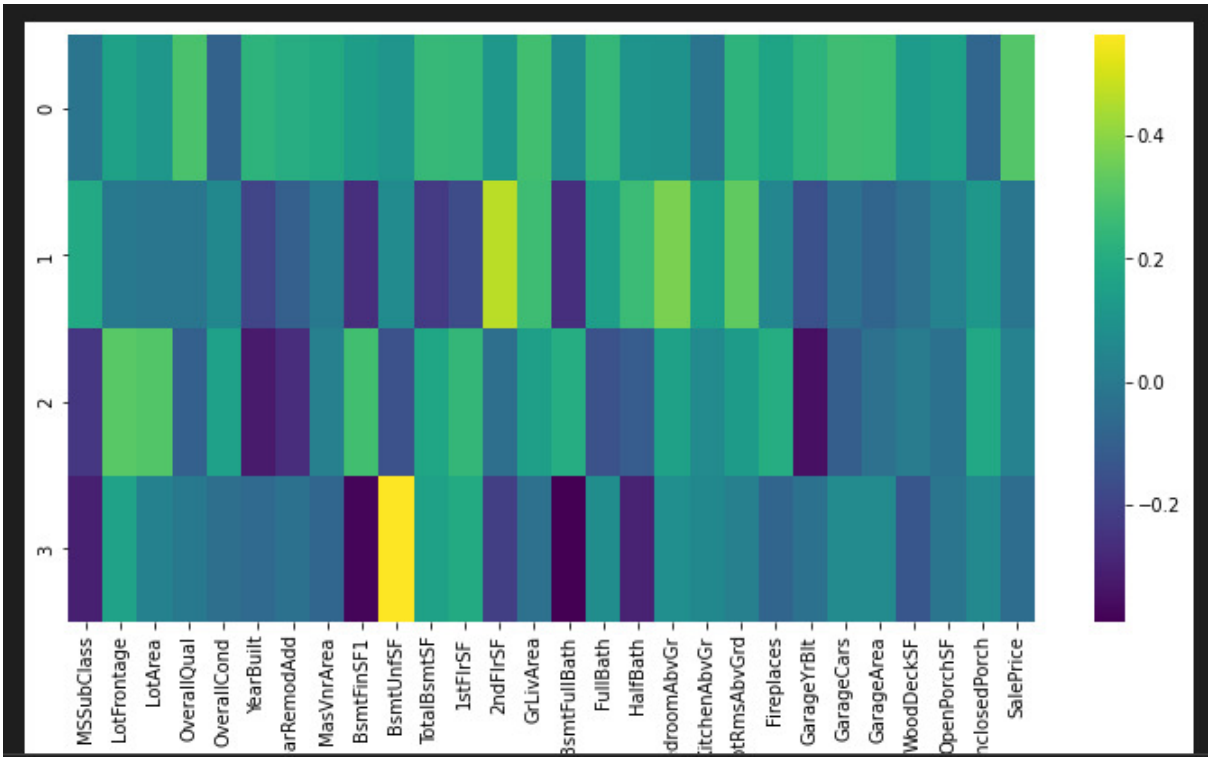
```
C:\Users\Oscar Paredez\AppData\Local\Temp\ipykernel_22852\1590771718.py:6:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  data_test.dropna(inplace=True)

(25992.46108530599, 0.0)
```

```
from factor_analyzer.factor_analyzer import calculate_kmo
kmo_all,kmo_model=calculate_kmo(data_test)
kmo_all,kmo_model
```

✓ 0.4s

```
(array([0.66511204, 0.87746015, 0.87616965, 0.946864  , 0.59567764,
        0.80266473, 0.85914329, 0.95438639, 0.59898192, 0.48276987,
        0.75302676, 0.6709207 , 0.53326766, 0.71708435, 0.86084578,
        0.91160162, 0.80352067, 0.79964645, 0.55242681, 0.92354268,
        0.9453402 , 0.87954873, 0.8867224 , 0.86602719, 0.95973055,
        0.94732983, 0.76345185, 0.94634604]),
 0.8014949962440353)
```





Se puede visualizar en el análisis de componentes principales que el test de esfericidad de Bartlett es válido para este conjunto de datos, ya que la matriz de correlaciones obtenida no representa una matriz de identidad, por lo que es un test adecuado en este caso.

Igualmente la técnica de análisis factorial es un test válido, como se puede observar en la matriz de KMO. En base a lo que se puede observar previamente, podemos decir que se seleccionan cuatro componentes principales para obtener la mayor variabilidad posible debido a que se pueden distinguir cuatro colores principales. Los coeficientes principales son las cuatro filas que se pueden observar en la matriz de PCA, las cuales representan de manera generalizada lo que representan todas las variables. En este caso, son cuatro componentes principales, por lo que son cuatro variables que logran generalizar el conjunto de datos.

En términos generales, se puede ver que el componente #0 explica la calidad de la casa, ya que engloba variables como el tamaño de la casa, la calidad de la misma, el precio de la casa, etc. El componente #1 resume aspectos del segundo piso de las viviendas, como el terreno que este segundo piso ocupa, la cantidad de cuartos en el segundo piso, etc. El componente #2 resume el área del terreno de la propiedad. Finalmente el componente #3 relaciona variables sobre el primer nivel de la casa.

7.

Sobre las reglas de asociación, se puede determinar que existen 1460 reglas de asociación. Las reglas más interesantes son por ejemplo las casas con AllPub (casas que incluyen los servicios de agua, luz, etc.) Se relacionan con las casas IR1, o sea que son casas irregulares. Otra regla interesante es que las casas que se encuentran conectadas a un solo camino tienen los servicios de AllPub. Así como estas dos reglas, existen muchas otras más entre las 1460 interesantes.

```
['Lvl']
Support: 0.8904109589041096
Confidence: 0.8904109589041096
['Pave']
Support: 0.9931506849315068
Confidence: 0.9931506849315068
['nan']
Support: 0.910958904109589
Confidence: 0.910958904109589
['AllPub', 'Corner']
Support: 0.136986301369863
Confidence: 1.0
['AllPub', 'IR1']
Support: 0.3013698630136986
Confidence: 1.0
['Inside', 'AllPub']
Support: 0.8013698630136986
Confidence: 0.8013698630136986
['AllPub', 'Lvl']
...
Confidence: 0.75
['Inside', 'Pave', 'AllPub', 'Reg', 'nan', 'Lvl']
Support: 0.4726027397260274
Confidence: 0.711340206185567
```

Hallazgos y Conclusiones

Como parte del análisis exploratorio, se puede concluir que es necesario realizar la separación de variables cuantitativas y cualitativas para proceder con el análisis exploratorio de maneras distintas. Asimismo, es importante mencionar que la herramienta de Profile Report es muy útil, ya que desarrolla por sí misma un análisis exploratorio bastante completo, y solamente necesita del conjunto de datos para poder desarrollar dicho análisis. En cuanto al valor de las variables se pudieron hacer limpieza de variables que no aportan ningún valor al conjunto de datos, permitiendo un análisis más efectivo, además se logró identificar relaciones que a futuro ayudarán a comprender los resultados otorgados por diferentes pruebas.

En cuanto al análisis de componentes principales, se puede decir que se seleccionan cuatro componentes principales para obtener la mayor variabilidad posible debido a que se pueden distinguir cuatro colores principales, además el mapa de correlación de los coeficientes principales hizo notar que estos se relacionan a términos generales de la casa, como por ejemplo uno de los que se lograron identificar fue en relación a la calidad del segundo piso y su área, otro calidad de la casa, el penúltimo sobre el terreno de la propiedad y el último en

relación al segundo piso, esto nos da mucho valor agregado para identificar puntos clave en nuestro conjunto de datos.

Sobre las reglas de asociación, se puede determinar que existen 1460 reglas de asociación. Las reglas más interesantes son por ejemplo las casas con AllPub (casas que incluyen los servicios de agua, luz, etc.) Se relacionan con las casas IR1, o sea que son casas irregulares. Otra regla interesante es que las casas que se encuentran conectadas a un solo camino tienen los servicios de AllPub.

Link del google drive:

https://docs.google.com/document/d/1xslP6_hrsfZQO89JfrMM_WqNCHqTPbC__RVtt9gac-s/edit?usp=sharing

Link del repo:

https://github.com/oscarparedez/labs-datascience/tree/main/lab1